Institute for Advanced Simulation

Two Topics in Ab Initio Molecular Dynamics: Multiple Length Scales and Exploration of Free-Energy Surfaces

Mark E. Tuckerman

published in

Multiscale Simulation Methods in Molecular Sciences,
J. Grotendorst, N. Attig, S. Blügel, D. Marx (Eds.),
Institute for Advanced Simulation, Forschungszentrum Jülich,
NIC Series, Vol. **42**, ISBN 978-3-9810843-8-2, pp. 177-201, 2009.

http://www.fz-juelich.de/nic-series/volume42

Two Topics in Ab Initio Molecular Dynamics: Multiple Length Scales and Exploration of Free-Energy Surfaces

Mark E. Tuckerman

Department of Chemistry and
Courant Institute of Mathematical Sciences
100 Washington Square East
New York University, New York, NY 10003
E-mail: mark.tuckerman@nyu.edu

This lecture will consider two problems that arise in molecular dynamics simulations of complex systems. The first is the treatment of multiple length scales in simulations that employ reciprocal-space techniques (plane-wave basis sets in *ab initio* molecular dynamics, Ewald summation,...) for the calculation of long-range forces. It will be shown that a dual-gridding scheme, with reciprocal space grids of two very different resolutions, can be used to substantially reduce the cost of the calculation without sacrificing accuracy. Interpolation between the two grids is achieved via the use of Euler exponential splines commonly employed in the particle-mesh Ewald method. Two application areas from *ab initio* molecular dynamics will be illustrated, namely, the use of dual-gridding in QM/MM calculations and in cluster calculations with plane-wave basis sets. The second problem is an inherently multiple time-scale problem involving the exploration of rough free-energy surfaces. It will be shown that efficient exploration and calculation of such surfaces is possible using an adiabatic dynamics technique in which a subset of collective variables are "driven" at high temperature by a set of external driving variables whose masses are adjusted so as to effect an adiabatic decoupling of the collective variables from the remainder of the system. Under these conditions, free-energy surfaces can be constructed straightforwardly from the probability distribution functions generated with the requirement of adjusting only a few parameters. The method will be illustrated on the folding of an alanine hexamer.

1 The Multiple Length-Scale Problem

Chemical systems are often characterized by a set of electronically active constituents localized in a small region of space, surrounded by and interacting with a large bath of electronically inert components. The division of a large system into chemically interesting and chemically uninteresting regions is both an intuitively appealing and a practically useful description that can be fruitfully applied to many different problems of chemical and biological interest. For example, in the study of solution phase chemical reactions, it is advantageous to consider, explicitly, the electronic degrees of freedom of the reactants and products and perhaps a first solvation shell, while the remainder of the solvent is modeled more approximately[1]. Similarly, in studies of enzyme catalysis, the valence electrons of the amino acids and the water molecules near the active site, as well as those of the substrate, must be modeled using a high level of theory while the remainder of these large and complex systems can be modeled more approximately[1–6]. Thus, simulation studies based on hybrid model descriptions promise to yield chemical insight into significant problems for low computational cost. It is, therefore, important to develop both the models and methods required to treat mixed *ab initio*/empirical force field descriptions of chemical and biological systems accurately and efficiently. In addition, systems with reduced peri-

odicity that require a vacuum region in the non-periodic direction, require methodological developments to reduce the inefficiency of treating a large empty spatial region.

It has been demonstrated that a wide variety of complex chemical systems can be treated effectively using an *ab initio* methodology that employs a plane wave basis set in conjunction with the generalized gradient approximation to density functional theory (GGA-DFT)[7–10]. Of course, in realistic calculations, many basis functions or equivalently, a large plane wave energy cutoff ($E_{cut} = \hbar^2 g_{max}^2/2m_e$) must be employed to ensure accuracy. The large basis set size coupled with the fact that plane waves, naturally, admit, only a single length scale has made it difficult to employ plane wave based GGA-DFT to study hybrid model systems.

Consider, for example, a small simulation cell containing the electron density embedded within a large simulation cell containing the rest of the system (i.e. the bath) or possibly empty space. In order to determine the long range interaction of the electron density with the atoms outside the small simulation cell within the plane wave formalism, it is necessary to expand the electron density in the large simulation cell using the **same** large cutoff required to describe the rapidly varying electron density in the small cell (e.g. $E_{cut} \approx 70$ Ry). Thus, the memory requirements are prohibitively large and the calculations scale poorly with the size of the large cell (at fixed small cell size). However, such a scheme does allow systems modeled using 3D periodic boundary conditions (liquids and solids) to be accurately studied. It also permits novel reciprocal space based techniques that treat clusters, wires and surfaces appropriately[11–13] (open, and 1D and 2D periodic boundary conditions, respectively), properly, to be applied to "mixed" or "hybrid" model calculations.

In this lecture, we will describe a dual-gridding method[14] designed to treat large systems that can be decomposed into electronically active and electronically inert portions with high efficiency is presented. Two length scales are explicitly introduced into the plane wave based GGA-DFT electronic structure formalism so that the small length scale, electronically active region can be treated differently than the long length scale, electronically inert region without loss of generality and with large gains in scalability and efficiency. This is accomplished by employing a Cardinal B-spline based formalism to derive a novel expression for the electronic energy that explicitly contains both the long and short length scales. The new expression can be evaluated efficiently using two independent plane wave energy cutoffs and is smooth, differentiable, and rapidly convergent with respect to the plane wave cutoff associated with the long length scale even when the plane wave cutoff associated with the short length scale is quite large. Thus, the method scales as $N \log N$ where N is number of atoms in the full system (at fixed size of the chemically active region) provided particle mesh Ewald techniques[15–18] are employed to evaluate the atomic charge density in the large cell. In addition, the new methodology does not involve an *ad hoc* electrostatic potential fitting scheme based on point charges derived from a particular choice of population analysis and can be utilized to treat clusters, wires, surfaces and solids/liquids without loss of generality. We note that a similar approach was recently developed for use in Gaussian-based QM/MM calculations[19].

1.1 Methods

In the Kohn-Sham formulation of density functional theory, the electron density is expanded in a set of orbitals $\{\psi_i(\mathbf{r})\}$

$$n(\mathbf{r}) = \sum_{i=1}^{n} |\psi_i(\mathbf{r})|^2 \tag{1}$$

and the energy functional is given by

$$E[n] = T_s[\{\psi_i\}] + E_{\rm H}[n] + E_{\rm xc}[n] + E_{\rm ext}[n] \tag{2}$$

where T_s is the kinetic energy of a system of noninteracting electrons, $E_{\rm H}$ is the Hartree energy, and $E_{\rm xc}$ is the exchange and correlation energy. For example, one could employ a Generalized Gradient Approximation such as the BLYP (Becke86 exchange and LYP correlation) functional[20,21]. In this case, Eq. (2) is referred to as a GGA-density functional.

In this work, the GGA-density functional, Eq. (2), is minimized by expanding the orbitals in a finite plane wave basis set and varying the expansion coefficients subject to the orthogonality constraints ($\langle\psi_j|\psi_i\rangle = \delta_{ij}$). The plane wave basis set is truncated by including all plane waves with kinetic energy less than or equal to a cutoff energy, $\hbar g^2/2m_e \leq E_{cut}$. Finally, core electrons, which are difficult to treat in a plane-wave basis set, are replaced by atomic pseudopotentials. Typical pseudopotentials contain a long range local contribution, $E_{\rm loc}$, and a short range angular momentum dependent nonlocal contribution, $E_{\rm nonloc}$ that serves to replace the core (i.e. $E_{\rm ext} = E_{\rm loc} + E_{\rm nonloc}$).

The GGA-density functional, therefore, contains only two terms that act at long range, specifically, the Hartree, $E_{\rm H}[n]$, and local pseudopotential, $E_{\rm loc}[n]$, energies defined by

$$E_{\rm H}[n] = \frac{e^2}{2} \sum_{\hat{\mathbf{S}}} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}' \, \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|} \tag{3}$$

$$E_{\rm loc}[n] = \sum_{\hat{\mathbf{S}}} \sum_{I=1}^{N} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \, \phi_{{\rm loc},I}(\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}) n(\mathbf{r}) \tag{4}$$

where $\mathbf{R}_I$ is the Cartesian position of the Ith ion, $\overleftrightarrow{\mathbf{h}}$ is the cell matrix whose columns contain the d cell vectors, $\det \overleftrightarrow{\mathbf{h}} = V$ is the volume, and $\hat{\mathbf{S}} = \{\hat{s}_a, \hat{s}_b, \hat{s}_c\}$ is a vector of integers indexing the periodic replicas (in clusters, only $\hat{\mathbf{S}} = \{0, 0, 0\}$ is allowed while in systems periodically replicated in three spatial dimensions, the three integers span the full range).

In plane-wave based calculations the orbitals and, the density are expanded as follows,[8,22]

$$\psi_j(\mathbf{r}) = \frac{1}{\sqrt{V}} \sum_{\hat{\mathbf{g}}} \bar{\psi}_j(\mathbf{g}) \exp(i\mathbf{g}\cdot\mathbf{r})$$

$$n(\mathbf{r}) = \frac{1}{V} \sum_{\hat{\mathbf{g}}} \bar{n}(\mathbf{g}) \exp(i\mathbf{g}\cdot\mathbf{r}), \tag{5}$$

where $\mathbf{g} = \overleftrightarrow{\mathbf{h}}^{-1} \hat{\mathbf{g}}$ and the vector of integers, $\hat{\mathbf{g}} = \{g_a, g_b, g_c\}$, indexes reciprocal space. Typically, a cutoff is introduced on the sums describing the orbitals such that $\hbar g^2/2m_e \leq$

E_{cut}. (Note, the reciprocal space summation for the density is over a reciprocal space defined by the appropriately larger cutoff, $E_{cut}^{(\mathrm{density})} = 4E_{cut}$.) It is convenient to express the Hartree and local external energies in reciprocal space

$$E_{\mathrm{H}} = \frac{e^2}{2V} \sum_{\hat{\mathbf{g}}}{}' |\bar{n}(\mathbf{g})|^2 \left[\frac{4\pi}{g^2} + \hat{\phi}^{(\mathrm{screen,Coul})}(\mathbf{g}) \right] + \left(\frac{e^2}{2V} \right) \hat{\phi}^{(\mathrm{screen,Coul})}(0) |\bar{n}(0)|^2 \tag{6}$$

$$E_{\mathrm{loc}} = \frac{1}{V} \sum_{\hat{\mathbf{g}}}{}' \sum_{I=1}^{N} \bar{n}^*(\mathbf{g}) \exp(-i\mathbf{g}\cdot\mathbf{R}_I) \left[\tilde{\phi}_{\mathrm{loc},I}(\mathbf{g}) - eq_I \hat{\phi}^{(\mathrm{screen,Coul})}(\mathbf{g}) \right] + \frac{1}{V} \sum_{I=1}^{N} \bar{n}(0) \left[\tilde{\phi}^{(0)}_{\mathrm{loc},I} - eq_I \hat{\phi}^{(\mathrm{screen,Coul})}(0) \right]. \tag{7}$$

Here, $\tilde{\phi}_{\mathrm{loc},I}$ denotes the Fourier Transform of the local pseudopotential, the prime indicates that the $\mathbf{g} = 0$ term is eliminated and the function, $\tilde{\phi}^{(0)}_{\mathrm{loc},I}$ is the non-singular part of the local pseudopotential at $\mathbf{g} = 0$. The screening function, $\hat{\phi}^{(\mathrm{screen,Coul})}(\mathbf{g})$, is added to treat systems with fewer than three periodic dimensions (clusters, surfaces, wires), as discussed in Refs.[11–13]. For a system periodically replicated in three spatial dimensions, it is identically zero.

It is clear that the standard expressions for the Hartree and local external energies given in Eq. (6) and Eq. (7), respectively, only possesses a single length scale. A second length scale can be introduced by first rewriting the real space expressions for these two energies using the identity $\mathrm{erf}(\alpha r) + \mathrm{erfc}(\alpha r) = 1$,

$$\begin{aligned} E_{\mathrm{H}}[n] &= \left\{ \frac{e^2}{2} \sum_{\hat{\mathbf{S}}} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}' \, \frac{n(\mathbf{r}) n(\mathbf{r}') \mathrm{erfc}(\alpha |\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|)}{|\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|} \right\} \\ &+ \left\{ \frac{e^2}{2} \sum_{\hat{\mathbf{S}}} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}' \, \frac{n(\mathbf{r}) n(\mathbf{r}') \mathrm{erf}(\alpha |\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|)}{|\mathbf{r} - \mathbf{r}' + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|} \right\} \\ &= E_{\mathrm{H}}^{(\mathrm{short})}[n] + E_{\mathrm{H}}^{(\mathrm{long})}[n] \end{aligned} \tag{8}$$

$$\begin{aligned} E_{\mathrm{loc}}[n] &= \left\{ \sum_{\hat{\mathbf{S}}} \sum_{I=1}^{N} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \; n(\mathbf{r}) \left[\phi_{\mathrm{loc},I}(\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}) + \frac{eq_I \mathrm{erf}(\alpha |\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|)}{|\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|} \right] \right\} \\ &- \left\{ \sum_{\hat{\mathbf{S}}} \sum_{I=1}^{N} \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \; n(\mathbf{r}) \left[\frac{eq_I \mathrm{erf}(\alpha |\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|)}{|\mathbf{r} - \mathbf{R}_I + \overleftrightarrow{\mathbf{h}} \hat{\mathbf{S}}|} \right] \right\} \\ &= E_{\mathrm{loc}}^{(\mathrm{short})}[n] + E_{\mathrm{loc}}^{(\mathrm{long})}[n]. \end{aligned} \tag{9}$$

Here, the first term in the curly brackets in each equation is short range while the second term is long range. Note, both $\phi_{\mathrm{loc},I}(\mathbf{r})$ and $-eq_I \mathrm{erf}(\alpha r)/r$ approach $-eq_I/r$, asymptotically where q_I is the charge on I^{th} ion core. In the limit $\alpha V^{1/3} >> 1$, the sum over

images in the first term of each expression (i.e. the short range parts) can be truncated at the first or nearest image with exponentially small error.

In order to proceed, it will be assumed that the electrons are localized in a particular region of the large cell described by $\overleftrightarrow{\mathbf{h}}$ which can be enclosed in a small cell, described by $\overleftrightarrow{\mathbf{h}}_s$, centered at the point, $\mathbf{R}_c$. That is, the orbitals and, hence, electron density are taken to vanish on the surface of $\overleftrightarrow{\mathbf{h}}_s$. Furthermore, it is assumed, for simplicity, that the a_s, b_s and c_s axes of $\overleftrightarrow{\mathbf{h}}_s$ are parallel to the a, b and c axes of $\overleftrightarrow{\mathbf{h}}$ such that $\overleftrightarrow{\mathbf{h}}^{-1}\overleftrightarrow{\mathbf{h}}_s = \overleftrightarrow{\mathbf{D}}$, a diagonal matrix. Thus, we can define,

$$\begin{aligned} \psi_j(\mathbf{r}_s + \mathbf{R}_c) &= \psi_{j,s}(\mathbf{r}_s) \\ n(\mathbf{r}_s + \mathbf{R}_c) &= n_s(\mathbf{r}_s) \end{aligned} \tag{10}$$

where, the $\mathbf{r}_s$ span the small cell and can be expressed as $\mathbf{r}_s = \overleftrightarrow{\mathbf{h}}_s\mathbf{s}$ with $0 \le s_\alpha \le 1$ and, both, $\psi_j(\mathbf{r}) \equiv 0$ and $n(\mathbf{r}) \equiv 0$ for $\mathbf{r}_s = \mathbf{r} - \mathbf{R}_c$ outside the domain of $\overleftrightarrow{\mathbf{h}}_s$. The orbitals and the electron density can be expanded in a plane wave basis set that spans the small cell, only,

$$\begin{aligned} \psi_{j,s}(\mathbf{r}_s) &= \frac{1}{\sqrt{V}_s}\sum_{\hat{\mathbf{g}}_s} \bar{\psi}_{j,s}(\mathbf{g}_s)\exp(i\mathbf{g}_s\cdot\mathbf{r}_s) \\ n_s(\mathbf{r}_s) &= \frac{1}{V_s}\sum_{\hat{\mathbf{g}}_s} \bar{n}_s(\mathbf{g}_s)\exp(i\mathbf{g}_s\cdot\mathbf{r}_s) \quad , \end{aligned} \tag{11}$$

where $\mathbf{g}_s = \overleftrightarrow{\mathbf{h}}_s^{-1}\hat{\mathbf{g}}_s$, the vector of integers, $\hat{\mathbf{g}}_s = \{g_{a,s}, g_{b,s}, g_{c,s}\}$, indexes the small reciprocal space and $V_s = \det\overleftrightarrow{\mathbf{h}}_s$ is the volume of the small cell. The plane wave energy cutoff is taken to be $E_{cut}^{(\mathrm{short})}$ (with the cutoff on the density $4E_{cut}^{(\mathrm{short})}$).

Given that the electron density is localized in the small cell, the short range components of the Hartree and local pseudopotential energies can be evaluated straightforwardly,

$$\begin{aligned} E_{\mathrm{H}}^{(\mathrm{short})}[n] &= \frac{e^2}{2}\int_{D(\overleftrightarrow{\mathbf{h}}_s)} d\mathbf{r}\int_{D(\overleftrightarrow{\mathbf{h}}_s)} d\mathbf{r}'\,\frac{n_s(\mathbf{r})n_s(\mathbf{r}')\mathrm{erfc}(\alpha|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|} \\ &= \frac{e^2}{2V_s}{\sum_{\hat{\mathbf{g}}_s}}'\bar{n}_s(-\mathbf{g}_s)\bar{n}_s(\mathbf{g}_s)\left[\frac{4\pi}{g_s^2}\right]\left[1-\exp\left(-\frac{g_s^2}{4\alpha^2}\right)\right] + \frac{e^2\pi}{2V_s\alpha^2}|n_s(0)|^2 \end{aligned} \tag{12}$$

$$\begin{aligned} E_{\mathrm{loc}}^{(\mathrm{short})}[n] &= \sum_{J=1}^{N_s}\int_{D(\overleftrightarrow{\mathbf{h}}_s)} d\mathbf{r}\; n_s(\mathbf{r})\left[\phi_{\mathrm{loc},J}(\mathbf{r}-\mathbf{R}_J+\mathbf{R}_c) + \frac{eq_J\mathrm{erf}(\alpha|\mathbf{r}-\mathbf{R}_J+\mathbf{R}_c|)}{|\mathbf{r}-\mathbf{R}_J+\mathbf{R}_c|}\right] \\ &= \frac{1}{V_s}{\sum_{\hat{\mathbf{g}}_s}}'\sum_{J=1}^{N_s}\bar{n}_s^*(\mathbf{g}_s)\exp(-i\mathbf{g}_s\cdot[\mathbf{R}_J-\mathbf{R}_c]) \\ &\times\left[\tilde{\phi}_{\mathrm{loc},J}(\mathbf{g}_s) + \frac{4\pi eq_J}{g_s^2}\exp\left(-\frac{g_s^2}{4\alpha^2}\right)\right] \\ &+ \frac{1}{V_s}\sum_{J=1}^{N_s}\bar{n}_s(0)\left[\tilde{\phi}_{\mathrm{loc},J}^{(0)} - \frac{eq_J\pi}{\alpha^2}\right]. \end{aligned} \tag{13}$$

where the J sum runs over the N_s ions within the small cell, the $\hat{\mathbf{g}}_s$ sum runs over the large reciprocal-space grid of the small cell and $\mathbf{R}_c$ is the position of the small cell inside the large. Since the full system is not periodic on $\overleftrightarrow{\mathbf{h}}_s$ but on $\overleftrightarrow{\mathbf{h}}$, Eqs. (12-13) will only yield the correct short range energy if $\alpha V_s^{1/3} >> 1$ and $n(\mathbf{r}_s)$ vanishes on the small cell boundary. The non-local pseudopotential energy is short range and is assumed to be evaluated within the small cell (only, considering the N_s ions in the small cell and using the small cell reciprocal space). Similarly, the exchange correlation and the electronic kinetic energies can also be evaluated in the small cell using standard techniques.

Next, the expressions for the long range portions of the Hartree and local pseudopotential energies must be formulated. This can be accomplished by expanding the electron density localized in the small cell in terms of the plane waves of the large cell. This expansion is permitted because the electron density, localized in the small cell, obeys periodic boundary conditions in the large cell (i.e. it is zero on the surface of $\overleftrightarrow{\mathbf{h}}$). Thus,

$$\begin{aligned} E_{\rm H}^{\rm (long)}[n] &= \frac{e^2}{2}\sum_{\hat{\mathbf{S}}}\int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}\int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}'\ \frac{n(\mathbf{r})n(\mathbf{r}')\mathrm{erf}(\alpha|\mathbf{r}-\mathbf{r}'+\overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{r}'+\overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|} \\ &= \frac{e^2}{2V}{\sum_{\hat{\mathbf{g}}}}'\bar{n}(-\mathbf{g})n(\mathbf{g})\left[\frac{4\pi}{g^2}\exp\left(-\frac{g^2}{4\alpha^2}\right)+\hat{\phi}^{(\rm screen,Coul)}(\mathbf{g})\right] \\ &+ \left(\frac{e^2}{2V}\right)\left[\hat{\phi}^{(\rm screen,Coul)}(0)-\frac{\pi}{\alpha^2}\right]|n(0)|^2 \end{aligned} \tag{14}$$

$$\begin{aligned} E_{\rm loc}^{\rm (long)}[n] &= -\sum_{\hat{\mathbf{S}}}\sum_{I=1}^{N}\int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}\ n(\mathbf{r})\left[\frac{eq_I\mathrm{erf}(\alpha|\mathbf{r}-\mathbf{R}_I+\overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|)}{|\mathbf{r}-\mathbf{R}_I+\overleftrightarrow{\mathbf{h}}\hat{\mathbf{S}}|}\right] \\ &= -\frac{e}{V}{\sum_{\hat{\mathbf{g}}}}'\bar{n}^*(\mathbf{g})S(\mathbf{g})\left[\frac{4\pi}{g^2}\exp\left(-\frac{g^2}{4\alpha^2}\right)+\hat{\phi}^{(\rm screen,Coul)}(\mathbf{g})\right] \\ &- \frac{e}{V}\bar{n}_s(0)S(0)\left[\hat{\phi}^{(\rm screen,Coul)}(0)-\frac{\pi}{\alpha^2}\right]. \end{aligned} \tag{15}$$

where

$$S(\mathbf{g}) = \sum_I q_I\exp(i\mathbf{g}\cdot\mathbf{R}_I) \tag{16}$$

is the atomic structure factor and

$$\begin{aligned} \bar{n}(\mathbf{g}) &= \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r}\exp[-i\mathbf{g}\cdot\mathbf{r}]n(\mathbf{r}) \\ &= \int_{D(\overleftrightarrow{\mathbf{h}_s})} d\mathbf{r}_s\exp[-i\mathbf{g}\cdot\mathbf{r}_s]n(\mathbf{r}_s+\mathbf{R}_c) \\ &= \int_{D(\overleftrightarrow{\mathbf{h}_s})} d\mathbf{r}_s\exp[-i\mathbf{g}\cdot(\mathbf{r}_s-\mathbf{R}_c)]n_s(\mathbf{r}_s) \end{aligned} \tag{17}$$

are the plane wave expansion coefficients of the electron density in the reciprocal space of the large cell, $\mathbf{g} = \overleftrightarrow{\mathbf{h}}^{-1}\ \hat{\mathbf{g}}$. The integral in Eq. (17) can be extended to cover the domain

described by the large cell without loss of generality because $n(\mathbf{r}_s + \mathbf{R}_c) \equiv 0$ outside of the small cell. Note, $\bar{n}(\mathbf{g}) = \bar{n}_s(\mathbf{g}_s)$ if $\overleftrightarrow{\mathbf{h}}_s \equiv \overleftrightarrow{\mathbf{h}}$ and $\mathbf{R}_c = 0$. Methods for the efficient evaluation of Eq. (17) and, hence, Eq. (14) and Eq. (15) are developed below.

First, it is clear from the long range/short range decomposition of the Hartree and local pseudopotential energies that a different plane wave cutoff can be introduced to treat each part. That is, one cutoff, $E_{cut}^{(\mathrm{short})}$, can be used to evaluate the short range components of the energy, Eq. (12) and Eq(13), and another, $E_{cut}^{(\mathrm{long})}$ can be used to evaluate the long range components, Eq. (14) and Eq.(15). While the long range/short range decomposition is general, it is expected that the short range contributions will be obtained by integration over functions that rapidly vary spatially while the long range contributions will be obtained by integration over a slowly varying function. Therefore, the short range energy contributions must be evaluated using a large reciprocal space cutoff (i.e. the standard $E_{cut}^{(\mathrm{density,short})} = 4E_{cut}^{(\mathrm{short})}$). In contrast, the long range part can be evaluated in reciprocal space using a small cutoff, $E_{cut}^{(\mathrm{long})} << E_{cut}^{(\mathrm{short})}$. Thus, by splitting the electronic energy into two parts, large gains in efficiency are possible.

Next, consider the case that the number of particles in the small cell, N_s and the small cell volume, V_s), are much less than their large cell counterparts ($N_s << N$ and $V_s << V$) as would be the case for a large, chemically inert bath surrounding a chemically active subsystem. The computational cost of evaluating the short range local pseudopotential and short range Hartree, exchange correlation, non-local pseudopotential and the electronic kinetic energy as well as the overlap matrix, $\langle\psi_{j,s}|\psi_{i,s}\rangle$, scales like $\sim N_s^3$. The computational cost of evaluating the long range part of the Hartree and local pseudopotential energies depends on the computational cost of evaluating the atomic charge density, S(**g**), and the plane wave expansion of the density in the large cell (see Eq. (17)). Since the atomic charge density can be evaluated in $N\log N$ using Particle Mesh Ewald techniques[15–17], if Eq. (17) could also be evaluated in $N\log N$, the computational cost of the method would then be $N\log N$ at fixed $\overleftrightarrow{\mathbf{h}}_s$ and N_s. (The present approach yields a linear scaling method because, at fixed particle density and plane wave cutoff, the number of plane waves increases linearly with particle number).

In order to achieve linear scaling, the electron density must be interpolated from the small cell where it is described by a plane wave expansion with a large cutoff, $E_{cut}^{(\mathrm{short})}$, to the large cell where it is described by a plane wave expansion with a small cutoff, $E_{cut}^{(\mathrm{long})}$, effectively. First, consider the Fourier components of the density

$$\bar{n}(\mathbf{g}) = \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \exp[-i\mathbf{g}\cdot\mathbf{r}] n(\mathbf{r}). \tag{18}$$

If $n(\mathbf{r})$ can be expressed in a finite plane wave basis,

$$n(\mathbf{r}) \equiv \frac{1}{V} \sum_{\hat{g}_a=-P_a/2+1}^{P_a/2} \sum_{\hat{g}_b=-P_b/2+1}^{P_b/2} \sum_{\hat{g}_c=-P_c/2+1}^{P_c/2} \exp(i\mathbf{g}\cdot\mathbf{r})\bar{n}(\mathbf{g}), \tag{19}$$

then the Fourier coefficients can also be determined (exactly) from a discrete sum over a

real space grid

$$\bar{n}(\mathbf{g}) \equiv \frac{V}{P_a P_b P_c} \sum_{\hat{s}_a=0}^{P_a-1} \sum_{\hat{s}_b}^{P_b-1} \sum_{\hat{s}_c=0}^{P_c-1} e^{-2\pi i \hat{g}_a \hat{s}_a/P_a} e^{-2\pi i \hat{g}_b \hat{s}_b/P_b} e^{-2\pi i \hat{g}_c \hat{s}_c/P_c} n(\overleftrightarrow{\mathbf{hs}}) \tag{20}$$

Here, P_a, P_b, and P_c are both the number of reciprocal lattice points along each direction and the number of points discretizing the $\mathbf{a}, \mathbf{b}, \mathbf{c}$ axes of the cell, and $s_\alpha = \hat{s}_\alpha / P_\alpha$. Importantly, Eq. (20) and its inverse, Eq. (19), can be evaluated using a three dimensional Fast Fourier Transforms (3D-FFT) in order $N \log N$. A spherical cutoff is introduced in reciprocal space by simply assuming that $n(\mathbf{r})$ is described by a basis in which $\bar{n}(\mathbf{g}) \equiv 0$ when $\hbar^2 |\mathbf{g}|^2 / 2m_e > E_{cut}$.

Next, consider a function, $f(\mathbf{r})$ with plane wave expansion coefficients,

$$\begin{aligned} \bar{f}(\mathbf{g}) &= \int_{D(\overleftrightarrow{\mathbf{h}})} d\mathbf{r} \ \exp[-i\mathbf{g}\cdot\mathbf{r}] f(\mathbf{r}) \\ &= V \int_0^1 ds_a \int_0^1 ds_b \int_0^1 ds_c e^{-2\pi i \hat{g}_a s_a} e^{-2\pi i \hat{g}_b s_b} e^{-2\pi i \hat{g}_c s_c} f(\overleftrightarrow{\mathbf{hs}}). \end{aligned} \tag{21}$$

that can be described on a finite reciprocal space (cf. Eq. (20)). In order to express the plane wave expansion coefficients, accurately, in terms of a sum over an arbitrary set of equally spaced discrete points in real space (as opposed to the continuous integrals given in Eq. (21) or the discretization required by Eq. (20)), it useful to introduce the Euler exponential spline

$$\exp\left(\frac{2\pi i \hat{g}_\alpha u}{\tilde{P}_\alpha}\right) = d_m(\hat{g}_\alpha, \tilde{P}_\alpha) \sum_{\hat{s}=-\infty}^{\infty} M_m(u-\hat{s}) \exp\left(\frac{2\pi i \hat{g}_\alpha \hat{s}}{\tilde{P}_\alpha}\right) + \mathcal{O}\left(\frac{2|\hat{g}_\alpha|}{\tilde{P}_\alpha}\right)^m$$

$$d_m(\hat{g}_\alpha, \tilde{P}_\alpha) = \frac{\exp\left(2\pi i (m-1)/\tilde{P}_\alpha\right)}{\left[\sum_{j=0}^{m-2} M_m(j+1) \exp\left(2\pi i \hat{g}_\alpha j / \tilde{P}_\alpha\right)\right]} \tag{22}$$

where $\hat{s}$ is an integer, u is a real number, m is the spline order assumed to be even and the $M_m(u)$ are the Cardinal B splines

$$M_2(u) = 1 - |u-1| \tag{23}$$

$$\begin{aligned} M_m(u) &= \left[\frac{u}{m-1}\right] M_{m-1}(u) + \left[\frac{m-u}{m-1}\right] M_{m-1}(u-1) \\ M_m(u) &\neq 0 \qquad\qquad 0 < u < m \\ M_m(u) &= 0 \qquad\qquad u \leq 0, u \geq m \end{aligned} \tag{24}$$

The Cardinal B splines satisfy the following sum rule and recursion relation:

$$\sum_{\hat{s}=\infty}^{\infty} M_m(u-\hat{s}) = 1$$

$$\frac{dM_m(u)}{du} = M_{m-1}(u) - M_{m-1}(u-1)$$

Inserting the Euler exponential spline into Eq. (21) yields a well defined approximation to $\bar{f}(\mathbf{g})$,

$$\bar{f}(\mathbf{g}) \approx \left[V d_m^*(\hat{g}_a, \tilde{P}_a) d_m^*(\hat{g}_b, \tilde{P}_b) d_m^*(\hat{g}_c, \tilde{P}_c) \right] \times \sum_{\hat{s}_a=0}^{\tilde{P}_a-1} \sum_{\hat{s}_b=0}^{\tilde{P}_b-1} \sum_{\hat{s}_c=0}^{\tilde{P}_c-1} e^{-2\pi i \hat{g}_a \hat{s}_a/\tilde{P}_a} e^{-2\pi i \hat{g}_b \hat{s}_b/\tilde{P}_b} e^{-2\pi i \hat{g}_c \hat{s}_c/\tilde{P}_c} f^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s}) \tag{25}$$

where

$$f^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s}) = \int_0^1 ds'_a \int_0^1 ds'_b \int_0^1 ds'_c \sum_{k_a=-\infty}^{\infty} \sum_{k_b=-\infty}^{\infty} \sum_{k_c=-\infty}^{\infty} f(\overleftrightarrow{\mathbf{h}}\mathbf{s}') \times M_m([s'_a - k_a]\tilde{P}_a - \hat{s}_a) M_m([s'_b - k_b]\tilde{P}_b - \hat{s}_b) M_m([s'_c - k_c]\tilde{P}_c - \hat{s}_c). \tag{26}$$

is the interpolation of $f(\mathbf{r})$ onto the discrete real space grid defined by $s_\alpha = \hat{s}_\alpha/\tilde{P}_\alpha$ and $0 \le \hat{s}_\alpha \le \tilde{P}_\alpha - 1$.

Equation (25) can be evaluated using a 3D-FFT in order $N \log N$ provided the function, $f^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s})$, defined on the discrete real space, can be constructed in a computationally efficient manner. In addition, Eq. (25) is smooth and possesses $m-2$ continuous derivatives. Note, if $\tilde{P}_a > m+1$ then each point in the continuous space, $\{s'_a, s'_b, s'_c\}$, is mapped to m^3 unique points on the discrete grid indexed by $\{\hat{s}_a, \hat{s}_b, \hat{s}_c\}$ due to the finite support of the $M_m(p)$ (see Eq. (23)). Also, it is important to choose $\tilde{P}_\alpha > P_\alpha$ to reduce the error inherent in the interpolation (see Eq. (22)).

It is now a simple matter to generate a computationally efficient and well defined approximation to the Fourier coefficients, $\bar{n}(\mathbf{g})$, of an electron density $n(\mathbf{r})$ that is assumed to be nonzero only in the small cell described by $\overleftrightarrow{\mathbf{h}}_s$. First, given that $\bar{n}_s(\mathbf{g}_s)$, defined in Eq. (11), exists on a finite reciprocal space, the identity given in Eq. (20) holds. Thus, the discrete form of the density can be inserted into Eq. (26) and the integrals performed using trapezoidal rule integration with loss of generality to yield the desired interpolation from the small cell to the large cell,

$$n^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s}) = \left[\frac{V_s}{V}\right] \left[\frac{1}{P_{a,s}P_{c,s}P_{c,s}}\right] \sum_{\hat{s}'_a=0}^{P_{a,s}-1} \sum_{\hat{s}'_b=0}^{P_{b,s}-1} \sum_{\hat{s}'_c=0}^{P_{c,s}-1} \sum_{k_a=-\infty}^{\infty} \sum_{k_b=-\infty}^{\infty} \sum_{k_c=-\infty}^{\infty} n_s(\overleftrightarrow{\mathbf{h}}_s\mathbf{s}') \times M_m([s'_a + S_{a,s} - k_a]\tilde{P}_a - \hat{s}_a) M_m([s'_b + S_{b,s} - k_b]\tilde{P}_b - \hat{s}_b) \times M_m([s'_c + S_{c,s} - k_c]\tilde{P}_c - \hat{s}_c). \tag{27}$$

Here, $\{P_{a,s}, P_{b,s}, P_{c,s}\}$ are defined by the size of the small cell reciprocal space (through the plane wave energy cutoff, $E_{cut}^{(\mathrm{short})}$), $s'_\alpha = \hat{s}'_\alpha/P_{\alpha,s}$, $\mathbf{S}_s = \overleftrightarrow{\mathbf{h}}^{-1}\mathbf{R}_c$, and $V_s/V = \det \overleftrightarrow{\mathbf{D}}$ while the $\{\tilde{P}_a, \tilde{P}_b, \tilde{P}_c\}$ are defined by the size of the large cell reciprocal space (through the energy cutoff, $E_{cut}^{(\mathrm{long})}$).

The desired plane wave expansion of the density, $\bar{n}(\mathbf{g})$, is constructed by inserting $n^{(\mathrm{conv})}(\overleftrightarrow{\mathbf{h}}\mathbf{s})$ into Eq. (25) and performing a 3D-FFT. Note, in the limit, $\tilde{P}_a = P_{a,s}$, $\tilde{P}_b = P_{b,s}$, $\tilde{P}_c = P_{c,s}$ or $E_{cut}^{(\mathrm{short})} = E_{cut}^{(\mathrm{long})}$, and $\overleftrightarrow{\mathbf{h}} = \overleftrightarrow{\mathbf{h}}_s$, then $\bar{n}_s(\mathbf{g}_s) \equiv \bar{n}(\mathbf{g})$ because Eq. (20) is exact for a finite reciprocal space and the Euler exponential splines are

exact at the knots. Importantly, Eq. (27) can be evaluated in order $N_s m^3$ **and** the (dense) discrete real space grid spanning the small cell, $\overleftrightarrow{\mathbf{h}}_s$, and the (sparse) discrete real space grid spanning the large cell, $\overleftrightarrow{\mathbf{h}}$, need not be commensurate. In addition, the separable form of the $M_m(p)$, which is a consequence of the choice $\overleftrightarrow{\mathbf{h}}^{-1}\overleftrightarrow{\mathbf{h}}_s = \overleftrightarrow{\mathbf{D}}$, allows the required $M_m(p)$ to be evaluated independently in order $mN_s^{1/3}$. Thus, the overall computational cost of constructing $\bar{n}(\mathbf{g})$ is $N\log N$ (dominated by the FFT). Finally, the resulting $\bar{n}(\mathbf{g})$ (i.e. obtained by inserting Eq. (27) into Eq. (25)) is continuously differentiable with respect to the expansion coefficients of the orbitals, $\bar{\psi}_{j,s}(\mathbf{g}_s)$, defined in Eq. (11).

1.2 Use in cluster calculations

In any *ab initio* molecular dynamics scheme in which long-range forces are computed in reciprocal space, e.g. plane-waves, Gaussians, DVRs, systems with reduced periodicity, e.g., clusters, surfaces, wires, can be treated using the screening function methodology developed by Martyna and Tuckerman[11–13]. Typically, when using the screening function, however, the size of the box needs to be roughly twice the maximum distance between the two furthest atoms in the cluster, which makes the use of this methodology somewhat more expensive than other techniques. The dual-gridding technique above can be used to circumvent this problem. In order to use the dual-gridding scheme in the context of a cluster calculation, for example, one simply uses two boxes (see Fig. 1): The central box contains the cluster system and is chosen large enough to contain the cluster with a small buffer region. The outer box can be chosen quite large, at least twice as large as the furthest distance between two atoms in the cluster (but it can be larger as well). The coarse grid is then used to describe the **g**-space density in the large box, and the scheme outlined in the Methods section is used to compute the long-range energies. The Hartree energy, for example, would be computed as

$$E_{\rm H} = \frac{1}{2V_s}\sum_{\mathbf{g}_s} |\bar{n}(\mathbf{g}_s)|^2 \tilde{\phi}^{(\rm short)}(\mathbf{g}_s) + \frac{1}{2V}\sum_{\mathbf{g}\neq(\mathbf{0},\mathbf{0},\mathbf{0})} |\bar{n}(\mathbf{g})|^2 \bar{\phi}^{(\rm long)}(\mathbf{g}) \qquad (28)$$

1.3 Illustrative examples

As a first, simple illustrative example, consider a simple Gaussian electron density,

$$n(\mathbf{r}) = \left(\frac{\kappa^2}{\pi}\right)^{3/2} \exp(-\kappa^2 r^2). \qquad (29)$$

The interaction of this density with a point charge located an arbitrary distance, r_0, away from its center can be determined, analytically, $E_{ext} = \mathrm{erf}(\kappa r_0)/r_0$. In Table I, the convergence of the total external energy to the analytical value is presented as a function of the large cell plane wave cutoff and Cardinal B-spline interpolation order, for various choices of r_0. The calculations were performed using the cluster boundary condition technique of reference[11] and fixed small cell plane wave cutoff ($E_{cut}^{(\rm short)} = 120$ Ry). In general, it can be seen that low B-spline interpolation orders and small plane wave cutoffs in the large

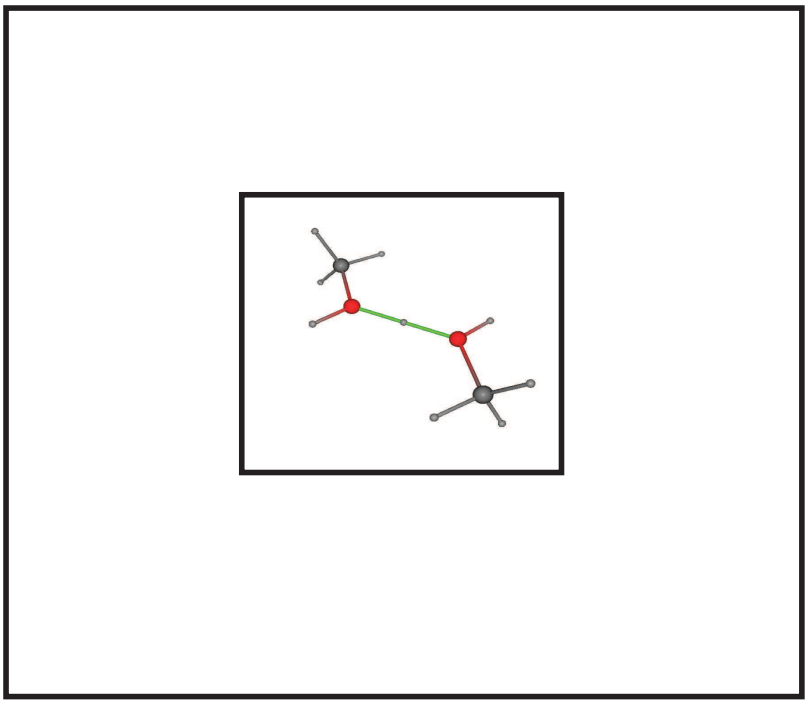

Figure 1. Illustration of the dual gridding scheme for cluster calculations.

cell, ($E_{cut}^{(\mathrm{long})}$), are sufficient to produce accurate results, indicating that the new method is both valid and efficient.

Next, we consider the human carbonic anhydrase II (HCA II) enzyme solvated in liquid water. In detail, the 260-residue HCA-II enzyme (complete with catalytic zinc – see Fig. 2), was solvated by 8,859 waters, for a total of 30,649 atoms. Clearly, a full *ab initio* treatment of such a large system is not feasible, at present. However, a hybrid model, wherein only the catalytic zinc, the side-chains of active site residues, HIS 94, HIS 96, HIS 119, THR 199, GLU 106 and the five water molecules in the active site are treated using an *ab initio* description, can be studied. Thus, 320 valence electrons of 80 atoms in the active site (see Fig. 2) are treated at an *ab initio* level while the remainder of the system is treated using the empirical CHARMM22 all-atom parameter force field which includes TIP3P water model[23]. Briefly, the electrons are assumed to interact with "*ab initio*" atoms via standard Troullier-Martins pseudopotentials[24] and with "empirical atoms" via pseudopotentials fit by the authors (see also[5,25]). The BLYP, density functional[20,21] was employed to treat exchange and correlation. *Ab initio* atoms (ion-cores) were permitted to interact with neighboring "empirical atoms" via appropriate bond, bend, torsion, one-four, van der Waals and Coulomb forces. The parameters were obtained by enforcing good agreement between mixed models, fully empirical models and fully *ab initio* models of relevant fragments. For example, the minimum energy geometry of hybrid model $\mathrm{CH_3CO-(HIS)-NHCH_3}$ deviates at most 2 degrees in the bend angles and 0.02Å in

Table 1. The interaction of a Gaussian charge density, $\kappa = 3.779454\mathring{A}^{-1}$, with a point charge at distance, r_0 from its center is presented as a function of large cell plane wave cutoff and B-Spline interpolation order. The large cell size was fixed at $L_l = 20\mathring{A}$ on edge. The small cell size was fixed at $L_s = 4\mathring{A}$ on edge and the small cutoff was fixed at $E_{cut}^{(\mathrm{short})} = 120Ry$. The electrostatic division parameter was set to be $\alpha = 6/L_s$ and $\Delta E_{ext} = E_{ext} - E_{ext}^{(exact)}$.

r_0 (Å)	$E_{cut}^{(\mathrm{long})}$ (Rydberg)	m	E_{ext} (Hartree)	ΔE_{ext} (Kelvin)
4	4	4	-0.132296	1
		6	-0.132297	1
		8	-0.132297	1
	8	4	-0.132293	0
		6	-0.132293	0
		8	-0.132293	0
6	4	4	-0.088186	3
		6	-0.088185	3
		8	-0.088185	3
	8	4	-0.088198	1
		6	-0.088198	1
		8	-0.088198	1
8	4	4	-0.066126	7
		6	-0.066125	7
		8	-0.066125	7
	8	4	-0.066149	1
		6	-0.066149	1
		8	-0.066149	1

the bond lengths from the standards (CHARMM and fully *ab initio* treatments as appropriate).

The HCA II/water system described above was prepared by taking the crystallographic configuration of the enzyme (PDB identification label, "1RAY")[26] and immersing it in TIP3P water. Next, a 1 ns constant temperature molecular dynamics calculation was performed using a fully empirical treatment[23]. This was followed by a 1 ns constant pressure molecule dynamics calculation. At this point, the hybrid model was introduced. In Table

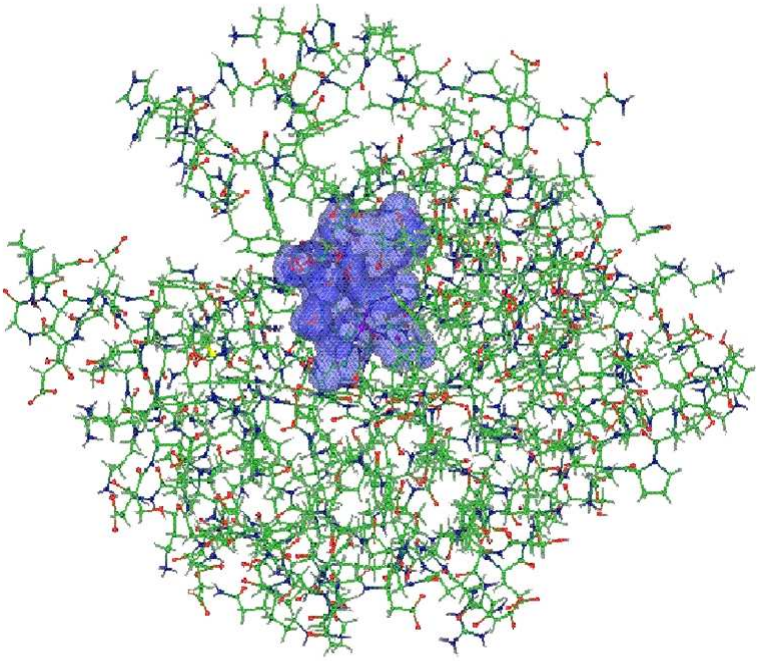

Figure 2. Snapshot of human carbonic anhydrase. The electronic density in the active site is shown as the blue isosurface.

Table 2. The total electronic energy of the active site of HCA II immersed in a bath of TIP3P molecules and CHARMM22 model amino acid residues as a function of large cell plane cutoff and spline interpolation order. The large cell size is fixed by the state point, 66.7Å, on edge. The small cell size was fixed at 18Å on edge and the small cell cutoff was fixed at 70Ry. The electrostatic division parameter was set to be $\alpha = 9/L_s$ and the accuracy measure is defined to be $\Delta E_{tot} = E_{tot}(E_{cut}^{(\mathrm{long})}, m) - E_{tot}(4, 8)$.

$E_{cut}^{(\mathrm{long})}$ (Rydberg)	m	E_{tot} (Hartree)	ΔE_{tot} (Kelvin)
0.5	6	-2329.31984	9200
	8	-2329.33018	5900
2	6	-2329.34896	32
	8	-2329.34905	3
4	6	-2329.34905	3
	8	-2329.34906	0

2, the convergence of the electronic energy for a representative configuration taken from the simulation of the hybrid model, is shown versus the large cell, plane wave cutoff and the Cardinal B-spline interpolation order. As is clear from the table, accurate energies are obtained for low spline orders and plane wave cutoffs.

2 Exploration of Free-Energy Surfaces

One of the key quantities in thermodynamics is the free energy associated with changes in conformation or thermodynamic state of a complex system. Molecular dynamics (MD) and Monte Carlo based approaches have emerged as important theoretical tools to study such free-energy changes along one-dimensional paths, e.g. along single reaction coordinates or one-dimensional λ-switching paths. Among these approaches, Umbrella Sampling[27–29] and Thermodynamic Integration[30–32] remain the most popular because they can be easily implemented. However, the problem of computing a multi-dimensional free-energy surface (FES) in several collective variables or reaction coordinates of interest has remained a significant challenge, particular when the FES contains numerous minima separated by high barriers. The mapping out of the free-energy landscape of small peptides and proteins in the Ramachandran angles, radius of gyration and/or number of hydrogen bonds, or the characterization of dissociation or mass-transfer processes in aqueous solution in terms of coordination numbers and distances are examples of this type of problem.

The challenge of treating such "rough" energy landscapes has lead to the introduction of various important new techniques for enhanced sampling of the configurational distribution of complex systems, from which the free energy is obtained. These include parallel tempering[33–38], hyperdynamics[39], parallel replica dynamics[40], Wang-Landau sampling[41,42], configuration-bias 'Monte Carlo[43], the Reference-Potential Spatial-Warping Algorithm[44,45], metadynamics[46], and techniques based on adiabatic dynamics[47–51], as a few examples. A comprehensive review of free-energy techniques was recently presented in the edited volume, *Free Energy Calculations*[52].

The Adiabatic Free Energy Dynamics (AFED),[47–49] introduced eight years ago by Rosso, *et al.*,[53] is a dynamical scheme for generating free-energy hypersurfaces in several collective variables of interest. The approach employs an imposed adiabatic decoupling between a small set of collective variables or reaction coordinates and the remaining degrees of freedom. Within this scheme, an elevated temperature is also applied to the collective variables to ensure that they are able to cross the high energy barriers needed to ensure sufficient sampling. In the limit of high temperature and adiabatic decoupling, it can be shown that the free energy hypersurface in the collective variables is obtained directly from their resultant probability distribution function[47,48]. The approach has been applied to the conformational sampling of small peptides[49] and in the computation of solvation and binding free energies via alchemical transformations[51]. In both cases, the use of adiabatic dynamics has been shown to lead to significant improvement in efficiency compared to traditional methods such as free-energy perturbation[54], umbrella sampling[27–29], and the blue moon ensemble approach[31,32]. In addition to being a relatively fast method, the AFED approach requires no *a posteriori* processing of the simulation data. Moreover, AFED is able to generate multi-dimensional free-energy hypersurfaces with significantly greater efficiency than multidimensional versions of the aforementioned approaches[49]. By construction, AFED generates full sweeps of the free-energy surface and, therefore, can

rapidly map out the locations of the free-energy minima well before the entire surface is fully converged.

The AFED approach is derived and implemented, in practice, by transforming the coordinate integrations in the canonical partition function to a set of generalized coordinates that explicitly contain the collective variables of interest. This gives rise to the disadvantage that the adiabatic dynamics must be carried out in these generalized coordinates, which leads to a steep implementation curve due to the rather invasive modifications to existing MD packages needed to introduce these transformations. It should be noted, however, that once such transformations are put in place, they can be subsequently combined with additional spatial-warping transformations that also significantly enhance conformational sampling[44,45].

Recently, Maragliano and Vanden-Eijnden[50] and, independently, Abrams and Tuckerman[55] built on the AFED approach by introducing a set of extended phase-space or "driving" variables that are harmonically coupled to the collective variables of interest. By imposing the adiabatic decoupling and high temperature on these extended variables rather than on the collective variables, the need for explicit transformations is avoided, thereby enlarging the class of collective variables that can be treated and rendering the technique substantially easier to implement. Maragliano and Vanden-Eijnden named the new technique "temperature accelerated molecular dynamics" or TAMD while Abrams and Tuckerman named if driven-AFED or d-AFED. It should be noted that such "driving" variables are also central in the so-called "metadynamics" approach[46], where they are used together with a time-dependent potential that floods energy basins with Gaussians, thereby allowing the system to escape the basin and move into a neighboring one. In metadynamics, as the basins are filled, the histogram in the collective variables becomes flat. When this occurs, the sum of all of the Gaussians is used to recover the free-energy hypersurface.

2.1 Adiabatic free-energy dynamics

Consider a system of N particles with Cartesian coordinates $\mathbf{r}_1, ..., \mathbf{r}_N \equiv \mathbf{r}$ and conjugate momenta $\mathbf{p}_1, ..., \mathbf{p}_N \equiv \mathbf{p}$ subject to a potential energy $V(\mathbf{r}_1, ..., \mathbf{r}_N)$. The classical canonical partition function for the system is given by

$$Q = C \int d^N\mathbf{p} \int_{D(V)} d^N\mathbf{r} \exp\left\{-\beta\left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V\left(\mathbf{r}_1, \ldots, \mathbf{r}_N\right)\right]\right\} \tag{30}$$

where $H(\mathbf{p}, \mathbf{r}) = \sum_i \mathbf{p}_i^2/2m_i + V(\mathbf{r})$ is the Hamiltonian, $\beta = 1/k_BT$, $D(V)$ is the spatial domain defined by the containing volume, and C is an overall prefactor that renders Q dimensionless and compensates for overcounting of states obtained by exchanging particles of the same chemical identity. For a system with M species, $C = [h^{3N} \prod_{\alpha=1}^{M} N_\alpha!]^{-1}$, where h is Planck's constant, and N_α is the number of particles of species α.

Suppose we wish to determine the free-energy hypersurface in a set of $n < N$ collective variables $q_1(\mathbf{r}), ..., q_n(\mathbf{r})$. Examples are the Ramachandran angles for characterizing the conformational space of oligopeptides or combinations of distances for characterizing a chemical reaction. The probability that $q_1(\mathbf{r})$ has the value s_1, $q_2(\mathbf{r})$ has the value s_2,...,$q_n(\mathbf{r})$ has the value s_n is given by

$$P(s_1, ..., s_n) = \frac{\int d^N\mathbf{p} d^N\mathbf{r}\ e^{-\beta H(\mathbf{p},\mathbf{r})} \prod_{i=1}^{n} \delta(q_i(\mathbf{r}) - s_i)}{\int d^N\mathbf{p} d^N\mathbf{r}\ e^{-\beta H(\mathbf{p},\mathbf{r})}} \tag{31}$$

Given this probability distribution, the free-energy hypersurface can be calculated according to

$$F(s_1, ..., s_n) = -kT \ln P(s_1, ..., s_n) \tag{32}$$

In many complex systems, direct calculation of the probability distribution function from a molecular dynamics trajectory is intractable because of the existence of high free-energy barriers separating important minima on the hypersurface. Free-energy surfaces of this type are said to be "rough", and it is necessary to employ enhanced sampling techniques. The adiabatic free-energy dynamics (AFED) achieves enhanced sampling in the variables $q_1(\mathbf{r}), ..., q_n(\mathbf{r})$ by introducing a high temperature $T_s \gg T$ for these n degrees of freedom only, while maintaining the remaining $3N - n$ degrees of freedom at the correct ensemble temperature T. The temperature disparity can be accomplished by introducing two separate sets of thermostats for each set of degrees of freedom. The high temperature T_s ensures that the variables $q_1, ..., q_n$ are able to cross high energy barriers on their part of the energy landscape. However, this high temperature also destroys the thermodynamic properties of the system *unless* the variables $q_1, ..., q_n$ are also adiabatically decoupled from the remaining degrees of freedom. In order to accomplish this decoupling, we need to be able to run the dynamics in a coordinate system that explicitly contains $q_1, ..., q_n$.

Suppose there is a transformation from Cartesian coordinates $\mathbf{r}_1, ..., \mathbf{r}_N$ to generalized coordinates $q_1, ..., q_{3N} \equiv q$ via the transformation equations $q_\alpha = q_\alpha(\mathbf{r})$. The inverse transformations are denoted $\mathbf{r}_i = \mathbf{r}_i(q)$. Substituting the transformation into Eq. (30) yields

$$\begin{aligned} Q &= C \int d^N \mathbf{p} \int_{D(V)} d^{3N} q \; J(q) \exp\left\{ -\beta \left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V\left(\mathbf{r}_1(q), ..., \mathbf{r}_N(q)\right) \right] \right\} \\ &= C \int d^N \mathbf{p} \int_{D(V)} d^{3N} q \exp\left\{ -\beta \left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \tilde{V}\left(q_1, ..., q_{3N}\right) \right] \right\} \end{aligned} \tag{33}$$

where the potential $\tilde{V}$ contains the Jacobian of the transformation $J(q) = |\partial \mathbf{r} / \partial q|$ and is given by

$$\tilde{V}(q_1, ..., q_{3N}) = V(\mathbf{r}_1(q), ..., \mathbf{r}_N(q)) - kT \ln J(q_1, .., q_{3N}) \tag{34}$$

Note that the partition function in Eq. (33) is completely equivalent to that in Eq. (30). Moreover, even though the transformation is not canonical, since we are only interested in sampling the ensemble distribution, we can treat the $3N$ Cartesian momentum components as "conjugate" to the $3N$ generalized coordinates, each being defined as $p_\alpha = m_\alpha \dot{q}_\alpha$, $\alpha = 1, ..., 3N$, where m_α are the associated masses. Thus, in order to achieve the desired adiabatic decoupling, we simply choose the first n masses m_α to be much larger than all of the remaining masses, $m_{1,...,n} \gg m_{n+1,...,3N}$.

Under the conditions of adiabatic decoupling and the temperature disparity, it was shown in Refs.[47,48], via a decomposition of the classical propagator, that the probability distribution, denoted $P_{\mathrm{adb}}(s_1, ..., s_n)$, becomes

$$P_{\mathrm{adb}}(s_1, ..., s_n) = \mathcal{N} \int d^n p \exp\left[-\beta_s \sum_{\alpha=1}^{n} \frac{p_\alpha^2}{2m_\alpha} \right] \left[Z(s_1, ..., s_n, \beta) \right]^{T/T_s} \tag{35}$$

where

$$Z(s_1,...,s_n,\beta) = \int d^{3N-n}p d^{3N}q \ \exp\left\{-\beta\left[\sum_{\alpha=n+1}^{3N}\frac{p_\alpha^2}{2m_\alpha} + \tilde{V}(q_1,...,q_{3N})\right]\right\} \times \prod_{\alpha=1}^{n}\delta(q_\alpha - s_\alpha) \tag{36}$$

and $\mathcal{N}$ is an overall normalization factor. In this case, because of the temperature ratio T/T_s in the exponent, the exact free-energy $F(s_1,...,s_n)$ at the temperature T, which is defined to be $F(s_1,...,s_n) = -kT\ln Z(s_1,...,s_n,\beta)$, is obtained from $P_{\rm adb}(s_1,...,s_n)$ by

$$F(s_1,...,s_n) = -kT_s \ln P_{\rm adb}(s_1,...,s_n) \tag{37}$$

Note that the multiplicative factor $-kT_s$ in Eq. (37) ensures that the free energy *at temperature* T is obtained. Eq. (37) shows that the free-energy surface can be computed *directly* from the probability distribution function generated in an adiabatic dynamics calculation. A detailed proof of the AFED method is given in Refs.[47,48].

2.2 Adiabatic free-energy dynamics without transformations

The AFED approach is a powerful one that is capable of generating multidimensional free-energy surfaces efficiently, as was shown in Refs.[48,49]. However, the need to work in generalized coordinates is a distinct disadvantage of the method, as this requires rather invasive modifications to existing molecular dynamics codes.

Recently, Maragliano and Vanden-Eijnden[50] and Abrams and Tuckerman[55] showed that AFED could be re-expressed in a set of extended phase-space variables in a manner similar to that used in the metadynamics approach of Laio and Parrinello[46], thereby circumventing the need for explicit coordinate transformations. This new formulation, which the authors called "Temperature Accelerated Molecular Dynamics" (TAMD) or "driven-AFED" (d-AFED) increases both the flexibility of the AFED method, allowing larger classes of collective variables to be treated, and the ease of implementation in existing packages.

TAMD/d-AFED can be derived as follows. We rewrite the product of δ-functions in Eq. (31) as the limit of a product of Gaussian functions[56]

$$\prod_{\alpha=1}^{n}\delta\left(q_\alpha(\mathbf{r}) - s_\alpha\right) = \lim_{\kappa\to\infty}\sqrt{\frac{\beta\kappa}{2\pi}}\exp\left[-\sum_{\alpha=1}^{n}\frac{\beta}{2}k\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2\right] \tag{38}$$

When Eq. (38) is substituted into Eq. (31), we obtain

$$P(s_1,...,s_n) = \lim_{\kappa\to\infty}\mathcal{N}_\kappa\int d^N\mathbf{p}\int_{D(V)} d^N\mathbf{r} \times \exp\left\{-\beta\left[H(\mathbf{p},\mathbf{r}) + \frac{1}{2}\kappa\sum_{\alpha=1}^{n}\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2\right]\right\} \tag{39}$$

where $\mathcal{N}_\kappa$ is a κ-dependent normalization constant. For large but finite κ, the integral in Eq. (39) represents a close approximation to the true probability distribution, and we

can regard the harmonic term in Eq. (39) as an additional potential term that keeps the collective variables $q_1(\mathbf{r}), ..., q_n(\mathbf{r})$ close to the values $s_1, ..., s_n$. In this representation, Eq. (39) resembles the probability distribution generated within the umbrella sampling approach[27–29]. However, if a set of n independent Gaussian integrations is introduced into Eq. (39) in the following form

$$P(s_1, ..., s_n) = \lim_{\kappa\to\infty} \mathcal{N}'_\kappa \int d^N\mathbf{p} \int_{D(V)} d^N\mathbf{r} \times \exp\left\{-\beta\left[H(\mathbf{p},\mathbf{r}) + \sum_{\alpha=1}^{n} \frac{p_{s_\alpha}^2}{2m_\alpha} + \frac{1}{2}\kappa\sum_{\alpha=1}^{n}\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2\right]\right\} \quad (40)$$

then the dependence of the distribution on $s_1, ..., s_n$ remains unaltered.

The argument of the exponential can now be regarded as an extended phase-space Hamiltonian

$$H_{\mathrm{ex}}(\mathbf{p}, p_s, \mathbf{r}, s) = \sum_{\alpha=1}^{n} \frac{p_{s_\alpha}^2}{2m_\alpha} + \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}_1, ..., \mathbf{r}_N) + \sum_{\alpha=1}^{n} \frac{1}{2}\kappa\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2 \quad (41)$$

This Hamiltonian generates the dynamics of the original N Cartesian positions and momenta and of the additional n variables $s_1, ..., s_n \equiv s$ and their conjugate momenta $p_{s_1}, ..., p_{s_n} \equiv p_s$ The extended variables serve to "drag" or "drive" the collective variables $q_1(\mathbf{r}), ..., q_n(\mathbf{r})$ via the harmonic coupling through their portion of the energy landscape provided that the variables $s_1, ..., s_n$ are able to sample a comparable region.

Assuming, again, that there are significant barriers hindering the sampling of the collective variables, enhanced sampling can be achieved by employing a high temperature and adiabatic decoupling, this time on the extended phase-space variables[50]. Thus, we introduce a temperature $T_s \gg T$ and masses $m_\alpha \gg m_i$ for these variables. As in the original AFED scheme, the former condition ensures that high barriers can be crossed, if T_s is chosen high enough, while the large masses ensure adiabatic decoupling of the extended phase-space variables from all other degrees of freedom. Following Refs.[47,48,51], it can be shown that, under these conditions, the distribution function generated takes the form

$$P_{\mathrm{adb}}^{(\kappa)}(s_1, ..., s_n) \propto \int d^n p \exp\left[-\beta_s \sum_{\alpha=1}^{n} \frac{p_{s_\alpha}^2}{2m_\alpha}\right] \left[Z(s_1, ..., s_n, \beta)\right]^{\beta_s/\beta} \quad (42)$$

where $\beta_s = 1/kT_s$ and

$$Z(s_1, .., s_n, \beta) = \int d^N\mathbf{p} \int_{D(V)} d^N\mathbf{r} \exp\left\{-\beta\left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \bar{V}(\mathbf{r}, s)\right]\right\} \quad (43)$$

and

$$\bar{V}(\mathbf{r}, s) = V(\mathbf{r}_1, ..., \mathbf{r}_N) + \frac{1}{2}\kappa\sum_{\alpha=1}^{n}\left(q_\alpha(\mathbf{r}) - s_\alpha\right)^2 \quad (44)$$

The probability distribution in Eq. (42) generates an approximation $F_\kappa(s_1, ..., s_n)$ to the true free energy profile at temperature T according to

$$F_\kappa(s_1, ..., s_n) = -kT_s \ln P_{\mathrm{adb}}^{(\kappa)}(s_1, ..., s_n) \quad (45)$$

and it is clear that in the limit $\kappa \to \infty$, the true free energy profile is recovered

$$F(s_1, ..., s_n) = \lim_{\kappa \to \infty} F_\kappa(s_1, ..., s_n) \tag{46}$$

Eqs. (45) and (46) show that the free-energy hypersurface can be generated within the adiabatic dynamics scheme without requiring a transformation to generalized coordinates.

The ability of TAMD and d-AFED to generate the free-energy surface efficiently depends on the thermostatting mechanism employed to maintain the two temperatures. The adiabatic decoupling represents a non-equilibrium steady state, and in Refs.[47,48], it was shown that the generalized Gaussian moment thermostat (GGMT) of Liu and Tuckerman[57] is an effective approach for maintaining the temperature disparity within the AFED scheme. Therefore, we employ it here as well. For completeness, we show the explicit equations of motion, including the coupling to separate GGMTs at temperatures T and T_s. As noted, GGMTs are capable of maintaining temperature control under the nonequilibrium (steady-state) conditions implied by the two temperatures and adiabatic decoupling. Within the two-moment version of the GGMT technique, with a separate thermostat coupled to each degree of freedom, the equations of motion for the d-AFED scheme read

$$\begin{aligned}
\dot{r}_{i,k} &= \frac{p_{i,k}}{m_i} \\
\dot{p}_{i,k} &= F_{i,k} - \kappa \sum_{\alpha=1}^{n} \left(q_\alpha(\mathbf{r}) - s_\alpha\right) \frac{\partial q_\alpha}{\partial r_{i,k}} - \frac{p_{\eta_{i,k,1}}}{Q_1} p_{i,k} - \frac{p_{\eta_{i,k,2}}}{Q_2} \left[(kT) p_{i,k} + \frac{p_{i,k}^3}{3m_i}\right] \\
\dot{s}_\alpha &= \frac{p_{s_\alpha}}{m_\alpha} \\
\dot{p}_{s_\alpha} &= \kappa \left(q_\alpha(\mathbf{r}) - s_\alpha\right) - \frac{p_{\xi_{\alpha,1}}}{Q_1'} p_{s_\alpha} - \frac{p_{\xi_{\alpha,2}}}{Q_2'} \left[(kT_s) p_{s_\alpha} + \frac{p_{s_\alpha}^3}{3m_\alpha}\right] \\
\dot{\eta}_{i,k,1} &= \frac{p_{\eta_{i,k,1}}}{Q_1} \\
\dot{\eta}_{i,k,2} &= \left[(kT) + \frac{p_{i,k,1}^2}{m_i}\right] \frac{p_{\eta_{i,k,2}}}{Q_2} \\
\dot{\xi}_{\alpha,1} &= \frac{p_{\xi_{\alpha,1}}}{Q_1'} \\
\dot{\xi}_{\alpha,2} &= \left[(kT_s) + \frac{p_{s_\alpha}^2}{m_\alpha}\right] \frac{p_{\xi_{\alpha,2}}}{Q_2'} \\
\dot{p}_{\eta_{i,k,1}} &= \frac{p_{i,k}^2}{m_i} - kT \\
\dot{p}_{\eta_{i,k,2}} &= \frac{p_{i,k}^4}{3m_i^2} - (kT)^2 \\
\dot{p}_{\xi_{\alpha,1}} &= \frac{p_{s_\alpha}^2}{m_\alpha} - kT_s \\
\dot{p}_{\xi_{\alpha,2}} &= \frac{p_{s_\alpha}^4}{3m_\alpha^2} - (kT_s)^2
\end{aligned} \tag{47}$$

where $F_{i,k} = -\partial V / \partial r_{i,k}$. In Eqs. (47), the thermostats are used to control the fluctuations in the second and fourth moments of the distribution of each momentum variable in the sys-

tem, whether these correspond to T or T_s. Here, k indexes the three Cartesian components of the physical coordinate and momentum of particle i, and $\eta_{i,k,1}$, $\eta_{i,k,2}$, $p_{\eta_{i,k,1}}$ and $p_{\eta_{i,k,2}}$ are the corresponding GGMT variables. Similarly, α indexes the extended phase-space driving variables, and $\xi_{\alpha,1}$, $\xi_{\alpha,2}$, $p_{\xi_{\alpha,1}}$ and $p_{\xi_{\alpha,2}}$ are the corresponding GGMT variables. The thermostat mass parameters Q_1, Q_2, Q_1', and Q_2' are chosen according to[57]:

$$\begin{aligned} Q_1 &= kT\tau^2 \qquad & Q_2 &= \frac{8}{3}(kT)^3\tau^2 \\ Q_1' &= kT_s\tau_s^2 \qquad & Q_2' &= \frac{8}{3}(kT_s)^3\tau_s^2 \end{aligned} \tag{48}$$

where τ and τ_s are characteristic time scales in the physical and extended systems, respectively. A typical choice for τ_s is the period of the harmonic coupling $(2\pi)\sqrt{m_\alpha/\kappa}$.

Another important feature of Eqs. (47) is that the presence of the stiff harmonic force term $(\kappa/2)\sum_\alpha (q_\alpha(\mathbf{r}) - s_\alpha)^2$ renders them amenable to multiple time-scale (r-RESPA) integration techniques[58,59]. For Eqs. (47), the Liouville operator, from which such integrators are derived, can be subdivided into a reference system containing the stiff oscillations of the harmonic coupling and a second propagator for the relatively slow motions associated with the motion of the physical system. Using r-RESPA to evaluate the inexpensive, but very fast, stiff harmonic force with a smaller timestep, while keeping the fundamental timestep larger, significantly improves the efficiency of the method. For details of this type of factorization the interested reader is directed to reference[58,59].

2.3 An illustrative example

The alanine hexamer (N-acetyl-(alanine)$_6$-methylamide) is a six-residue peptide that exhibits helical properties in solution[60]. Furthermore, computational studies of the alanine hexamer in solution, parameterized with the AMBER force field, results in helical conformational minima[61].

The simulation was performed by solvating the alanine hexamer in 698 TIP3P water molecules in a 27.9737 Å cubic periodic box. The molecule was started in its completely extended conformation and then equilibrated for 200 ps at constant volume (NVT), 1 ns at constant pressure (NPT), and finally 500 ps at constant volume (NVT).

In applying the d-AFED/TAMD method with the AMBER (parm94) force field, the two collective variables of interest chosen were the radius of gyration (R_G) and the number of intramolecular hydrogen bonds (N_H). These collective variables are defined as follows[62]

$$\begin{aligned} R_G &= \sqrt{\frac{1}{N_b}\sum_{i=1}^{N_b}\left(\mathbf{r}_i - \frac{1}{N_b}\sum_{j=1}^{N_b}\mathbf{r}_j\right)^2} \\ N_H &= \sum_{i=1}^{N_{\mathrm{Ox}}}\sum_{j=1}^{N_{\mathrm{Hy}}}\frac{1-\left(\frac{\mathbf{r}_i-\mathbf{r}_j}{d_0}\right)^6}{1-\left(\frac{\mathbf{r}_i-\mathbf{r}_j}{d_0}\right)^{12}} \end{aligned} \tag{49}$$

where N_b is the number of heavy backbone atoms, where N_{Ox} and N_{Hy} are the number of oxygen and hydrogen atoms, respectively, and $d_0 = 2.5$Å. Corresponding extended coordinates, s_1 and s_2, were added and these coordinates were treated as the slow variables

with masses $m_{s_1,s_2} = 15m_C$, where m_C is the mass of a carbon atom, and heated to a temperature of $T_{s_1,s_2} = 600$ K, while the physical variables were kept at a temperature of 300 K. s_1 and s_2 were coupled to R_G and N_H with harmonic coupling constants 5.4×10^6 K $\cdot$ Å^{-2} and 5.4×10^6 K, respectively.

The free energy surface computed from a relatively short production run of 5 ns, for the alanine hexamer in solution, leads to the free energy surface shown in Figure 3. The most significant feature of this surface is the presence of a global minimum at $(R_G, N_H) = (3.8, 4.4)$, corresponding to a right-handed helical, α_{R} , conformation (Figure 4a). Furthermore, there is also the presence of a large region of deformed/partial helices all with free energies lower than $+2$ kcal $\cdot$ mol^{-1} above the global minimum. These preliminary results are very consistent with previous results[61] for the alanine hexamer using the AMBER force field.

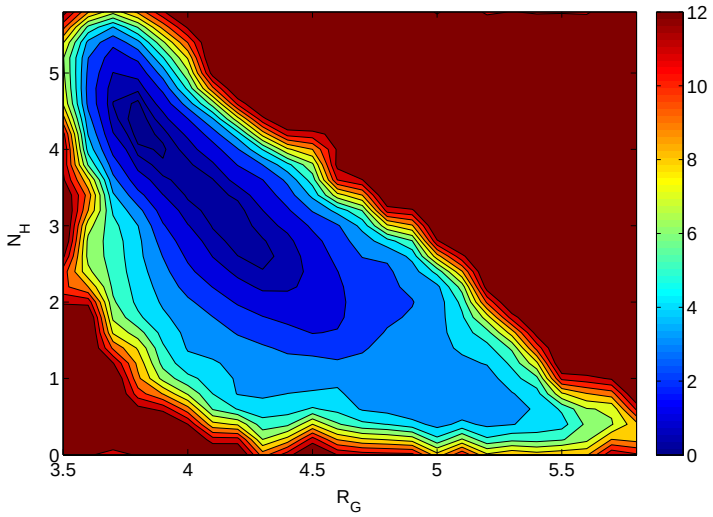

Figure 3. (a) Free-energy surface $F(R_G,\ N_H)$ and (b) contour plot computed for N-acetyl-(alanine)$_6$-methylamide (alanine hexamer) in solution with the AMBER force field.

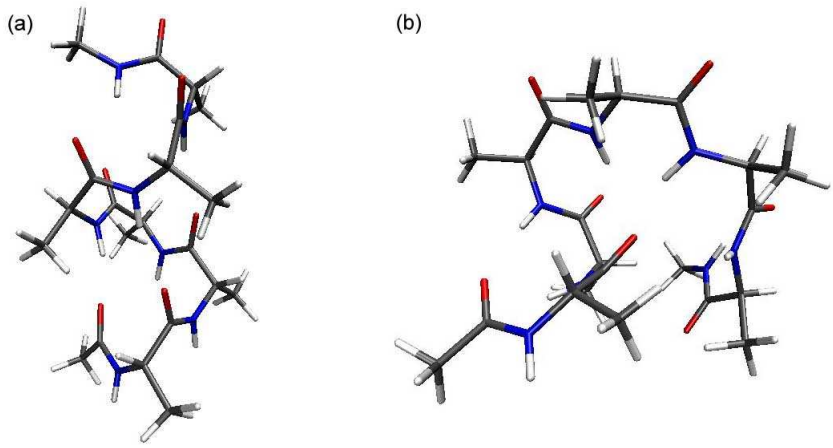

Figure 4. (a) Right-handed helical conformation of the alanine hexamer (global minimum). (b) Misfolded, extended, conformation of the alanine hexamer. All waters have been deleted for clarity.

References

1. A. Warshel and M. Levitt, *Theoretical studies of enzymic reactions – Dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme*, J. Mol. Biol., **103**, 227, 1976.
2. K. P. Eurenius, D. C. Chatfield, B. R. Brooks, and M. Hodoscek, *Enzyme mechanisms with hybrid quantum and molecular mechanical potentials: 1. Theoretical considerations*, Int. J. Quantum Chem., **60**, 1189, 1996.
3. Mulholland AJ, Lyne PD, and Karplus M, *Ab initio QM/MM study of the citrate synthase mechanism. A low-barrier hydrogen bond is not involved*, J. Am. Chem. Soc., **122**, 534, 2000.
4. A. Laio, J. VandeVondele, and U. Rothlisberger, *A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations*, J. Chem. Phys., **116**, 6941, 2002.
5. Y. Zhang, T. Lee, and W. Yang, *A pseudobond approach to combining quantum mechanical and molecular mechanical methods*, J. Chem. Phys., **110**, 46, 1999.
6. G. Alagona, P. Desmeules, C. Ghio, and P. A. Kollman, *Quantum mechanical and molecular mechanical studies on a model for the dihydroxyacetone phosphate blyceraldehyde phosphate isomerization catalyzed by triosephosphate isomerase (TIM)*, J. Am. Chem. Soc., **106**, 3623, 1984.
7. R. Car and M. Parrinello, *Unified approach for molecular dynamics and density functional theory*, Phys. Rev. Lett., **55**, 2471, 1985.
8. G. Galli and M. Parrinello, *The Ab-initio MD method*, Comp. Sim. Mat. Sci., **3**, 283, 1991.
9. D. Marx and J. Hutter, *Ab initio molecular dynamics: Theory and implementation*, Modern Methods and Algorithms for Quantum Chemistry, **1**, 301, 2000.

10. M. E. Tuckerman, *Ab initio molecular dynamics: Basic concepts, current trends and novel applications*, J. Phys. Condens. Matter, **14**, R1297, 2002.
11. G. J. Martyna and M. E. Tuckerman, *A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters*, J. Chem. Phys., **110**, 2810, 1999.
12. M. E. Tuckerman, P. Minary, K. Pihakari, and G. J. Martyna, *A new reciprocal space based treatment of long range forces on surfaces*, J. Chem. Phys., **116**, 5351, 2002.
13. P. Minary, J. A. Morrone, D. A. Yarne, M. E. Tuckerman, and G. J. Martyna, *Long range interactions on wires: A reciprocal space based formalism*, J. Chem. Phys., **121**, 11949, 2004.
14. D. Yarne, G. J. Martyna, and M. E. Tuckerman, *A Dual Space, Plane wave based approach to QM/MM calculations*, J. Chem. Phys., submitted (2001).
15. U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, *A smooth particle-mesh Ewald method*, J. Chem. Phys., **103**, 8577, 1995.
16. T.A. Darden, A. Toukmaji, and L.G. Pedersen, *Long-range electrostatic effects in biomolecular simulations*, J. Chim. Phys., **94**, 1346, 1997.
17. E.L. Pollock and J. Glosli, *Comments on P(3)M, FMM, and the Ewald method for large periodic coulombic systems*, Comp. Phys. Comm., **95**, 93, 1996.
18. M. Patra, M. Karttunen, M. T. Hyvonen, E. Falck, P. Lindqvist, and I. Vattulainen, *Molecular dynamics simulations of lipid bilayers: Major artifacts due to truncating electrostatic interactions*, Biophys. J., **84**, 3636, 2003.
19. T. Laino, F. Mohamed, A. Laio, and M. Parrinello, *An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations*, J. Chem. Theory. Comput., **2**, 1370, 2006.
20. A. D. Becke, *Density-functional exchange-energy approximation with correct asymptotic behavior*, Phys. Rev. A, **38**, 3098, 1988.
21. C. Lee, W. Yang, and R. G. Parr, *Development of the Colle-Salvetti correlation-energy formular into a functional of the electron density*, Phys. Rev. B, **37**, 785, 1988.
22. D. K. Remler and P. A. Madden, *Molecular dynamics without effective potentials via the Car-Parrinello approach*, Mol. Phys., **70**, 921–951, 1990.
23. A. D. MacKerell, D. Bashford, M. Bellott, Jr. R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, III W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B, **102**, 3586, 1998.
24. N. Troullier and J. L. Martins, *Efficient pseudopotentials for plane-wave calculations*, Phys. Rev. B, **43**, 1993, 1991.
25. J. Gao, P. Amara, and M. J. Field, *A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations*, J. Phys. Chem. A, **109**, 4714, 1998.
26. B. M. Jonsson, H. Hakansson, and A. Liljas, *The structure of human carbonic anhydrase-II in complex with bromide and azide*, FEBS Lett., **322**, 186, 1993.
27. G. M. Torrie and J. P. Valleau, Chem. Phys. Lett., **28**, 578, 1974.
28. G. M. Torrie and J. P. Valleau, J. Comput. Chem., **23**, 187, 1977.
29. B. J. Berne, (Ed.), *Modern Theoretical Chemistry V*, Plenum, New York, 1977.
30. J. G. Kirkwood, J. Chem. Phys., **3**, 300, 1935.

31. E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Chem. Phys. Lett., **156**, 472, 1989.
32. Michiel Sprik and Giovanni Ciccotti, *Free energy from constrained molecular dynamics*, J. Chem. Phys., **109**, 7737, 1998.
33. R. H. Swendsen and J. S. Wang, Phys. Rev. Lett., **57**, 2607, 1986.
34. K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn., **65**, 1604, 1996.
35. M. Tesi, E. J. J. Rensburg, E. Orlandini, and S. G. Whittington, J. Stat. Phys., **82**, 155, 1996.
36. P. Liu, B. Kim, R. A. Friesner, and B. J. Berne, Proc. Natl. Acad. Sci., **102**, 13749, 2005.
37. C. J. Woods, J. W. Essex, and M. A. King, *The development of replica-exchange-based free-energy methods*, J. Phys. Chem. B, **107**, 13703, 2003.
38. C. J. Woods, J. W. Essex, and M. A. King, *Enhanced Configurational Sampling in Binding Free-Energy Calculations*, J. Phys. Chem. B, **107**, 13711, 2003.
39. A. F. Voter, *Hyperdynamics: Accelerated molecular dynamics of infrequent events*, Phys. Rev. Lett., **78**, 3908, 1997.
40. A. F. Voter, *Parallel replica method for dynamics of infrequent events*, Phys. Rev. B, **57**, R13985, 1998.
41. F. G. Wang and D. P. Landau, *Efficient, multiple-range random walk algorithm to calculate the density of states*, Phys. Rev. Lett., **86**, 2050, 2001.
42. F. G. Wang and D. P. Landau, *Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram*, Phys. Rev. E, **64**, 056101, 2001.
43. J. I. Siepmann and D. Frenkel, *Configurational bias monte-carlo - a new sampling scheme for flexible chains*, Mol. Phys., **75**, 59, 1992.
44. Z. Zhu, M. E. Tuckerman, S. O. Samuelson, and G. J. Martyna, *Enhancing sampling of biomolecules using novel variable transformations*, Phys. Rev. Lett., **88**, 100201, 2002.
45. P. Minary, M. E. Tuckerman, and G. J. Martyna, *Dynamical spatial warping: A novel method for the conformational sampling of biophysical structure*, SIAM J. Sci. Comput., **30**, 2055, 2008.
46. A. Laio and M. Parrinello, *Escaping the free-energy minima*, Proc. Natl. Acad. Sci., **99**, 12562, 2002.
47. L. Rosso and M. E. Tuckerman, *An adiabatic molecular dynamics method for the calculation of free energy profiles*, Mol. Simul., **28**, 91, 2002.
48. L. Rosso, P. Minary, Z. Zhu, and M. E. Tuckerman, *On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles*, J. Chem. Phys., **116**, 4389, 2002.
49. L. Rosso, Jerry B. Abrams, and M. E. Tuckerman, *Mapping the backbone dihedral free-energy surfaces in small peptides in solution using adiabatic free-energy dynamics*, J. Phys. Chem. B, **109**, 4162–4167, 2005.
50. L. Maragliano and E. Vanden-Eijnden, *A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations*, Chem. Phys. Lett., **426**, 168, 2006.
51. Jerry B. Abrams, L. Rosso, and M. E. Tuckerman, *Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics*, J. Chem. Phys., **125**, 074115,

2006.
52. C. Chipot and A. Pohorille, editors, *Free Energy Calculations*, Springer, Heidelberg, 2007.
53. ", The technique was first presented at a CECAM meeting on free energy calculations held in 2000 from June 19-21. The proceedings of this meeting were published in *Molecular Simulation*, Volume 28, Issues 1-2 (2002), which contains Ref.[47].
54. R. W. Zwanzig, J. Chem. Phys., **22**, 1420, 1954.
55. J. B. Abrams and M. E. Tuckerman, *Efficient and direct generation of multidimensional free-energy surfaces via aidabatic dynamics without coordinate transformations*, J. Phys. Chem. B, **112**, 15742, 2008.
56. E. Butkov, *Mathematical Physics*, Addison-Wesley, Reading, 1968.
57. Yi Liu and Mark E. Tuckerman, *Generalized Gaussian moment thermostatting: A new continuous dynamical approach to the canonical ensemble*, J. Chem. Phys., **112**, 1685–1700, 2000.
58. M. Tuckerman, B. J. Berne, and G. J. Martyna, *Reversible multiple time scale molecular-dynamics*, J. Chem. Phys., **97**, 1990, 1992.
59. G. J. Martyna, Mark E. Tuckerman, D. J. Tobias, and M. L. Klein, Mol. Phys., **87**, 1117, 1996.
60. R. J. Kennedy, K. Y. Tsang, and D. S. Kemp, ", 2002.
61. N. Kamiya, Y. S. Watanabe, S. Ono, and J. Higo, ", 2005.
62. G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello, *Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics*, J. Am. Chem. Soc., **128**, 13435, 2006.